

METHOD AND APPARATUS FOR DETERMINING A SECOND  
PICTURE FOR TEMPORAL DIRECT-MODE BLOCK PREDICTION

Field of the Invention

5           The present invention relates to video generally and, more particularly, to a method and apparatus for determining a second picture for temporal direct-mode block prediction.

Background of the Invention

10           The H.264/MPEG4-AVC video standard allows multiple different reference pictures for inter-prediction. The different reference pictures are potentially signaled down to 8 x 8 partitions which reference a picture to use for inter-prediction. The standard also allows the choice, in a flexible manner, of which  
15 reference pictures to use, and the order in which the reference pictures, are available for any given slice (i.e., a group of macroblocks) of video.

          Such flexibility leaves the direct (i.e., spatial and temporal) block prediction modes open to a wide variety of  
20 different implementations. A direct-mode block is a bi-predictive

03-1431  
1496.00341

predicted block in a B-frame that does not signal either references or motion vectors. Rather, references and motion vectors are derived from a co-located block in a previously decoded picture. The overhead of the derived block mode is very low and provides a very important prediction mode that is often used to significantly reduce the rate of B-frames.

The reference pictures for each slice of video are arranged into two ordered lists (i.e., List0 and List1). For bi-predictive and direct-mode predicted blocks, one picture from each list should be indicated for use for inter-prediction by two reference-indices (one into each list) indicating an ordered number of one of the reference pictures from each list.

Previous H.264 implementations of direct-modes use the following sequence to determine which two current reference pictures should be used for inter-prediction of each block of direct-mode block. First, previous H.264 implementations find the co-located picture (i.e., reference 0, the first reference picture from List1) and block for the current block. This co-located picture will be the first reference picture used for direct-mode prediction. Next, the co-located block will be used to derive the reference indices and motion vectors for the current block.

03-1431  
1496.00341

Specifically, previous H.264 implementations determine the List0 reference picture that is used by the co-located block to refer to a 'direct-mode reference'. The reference index in the co-located picture of this direct mode reference is called the direct-mode reference index. The direct mode reference index is used by the current block to determine the second reference picture to use for inter prediction. Specifically, the direct-mode reference index is directly used in the reference picture list of the current slice. Finally, the motion vectors for the current block are interpolated from the motion vectors used in the co-located block according to the temporal distances between the current picture and the two reference pictures.

Such an implementation has the disadvantage that the second reference picture does not necessarily refer to the same physical reference picture for direct-prediction that was used for inter-prediction by the co-located block. The reference picture used in the co-located block and the second picture used in the direct-mode prediction of the current block are the same physical picture only if the direct mode reference picture was present in the same position (i) in List0 of the current slice of the current

03-1431  
1496.00341

picture being decoded and (ii) in List0 of the co-located slice of the co-located picture.

The intent of direct-mode prediction is that it uses the physical reference picture used by the co-located block as a reference picture for the current block. However, since H.264 supports reference picture re-ordering, this condition is not necessarily met. Reference picture re-ordering is the ability to flexibly order reference lists for each slice to use different pictures that are best inter-predicted from various other previously encoded/decoded pictures. If the encoder has the ability to specify which pictures are best for the current picture, then prediction residuals may be reduced.

A particular example of where the ability to re-order reference pictures is useful is to adaptively choose whether to code an I or P-picture as two fields (the second of which is inter-predicted from the first) or as a single picture without inter-prediction between fields. The reference pictures may be re-ordered between the current picture and the co-located picture such that the same reference picture does not occur in the same position in the respective List0. The direct-mode prediction could be seriously compromised with the existing solution since the intended

03-1431  
1496.00341

use of the direct-mode is that the same reference picture would be used.

It would be desirable to identify reference index that spatial and temporal direct-mode prediction modes should use to  
5 reference the picture that was the primary reference of the co-located macroblock.

#### Summary of the Invention

The present invention concerns a method for determining  
10 a first and a second reference picture used for inter-prediction of a macroblock, comprising the steps of (A) finding a co-located picture and block, (B) determining a reference index, (C) mapping the reference index to a lowest valued reference index in a current reference list and (D) using the reference index to determine the  
15 second reference picture.

The objects, features and advantages of the present invention include providing a method and/or apparatus that may (i) determine a second picture for temporal direct-mode block prediction and/or (ii) map a reference index to a lowest valued  
20 reference index in a current reference list.

**Brief Description of the Drawings**

These and other objects, features and advantages of the present invention will be apparent from the following detailed description and the appended claims and drawings in which:

5           FIG. 1 is a flow diagram of an implementation of the present invention;

          FIG. 2 is a diagram illustrating an implementation of the present invention;

          FIG. 3 is a partial block diagram of an example  
10       implementation of an encoder apparatus; and

          FIG. 4 is a partial block diagram of an example implementation of a decoder apparatus.

**Detailed Description of the Preferred Embodiments**

15           Referring to FIG. 1, a flow diagram 100 of the present invention is shown. The present invention may determine which two current reference pictures should be used for inter-prediction of each block of an H.264 direct-mode macroblock. The flow diagram  
100 generally comprises a state 102, a state 104, a state 106 and  
20       a state 108.

03-1431  
1496.00341

The state 102 finds the co-located (e.g., reference 0, a first reference picture from List1) picture and block for the current block being processed. The co-located picture will be the first reference picture used for direct-mode prediction. The state 5 104 determines the reference picture that was used by the co-located block to refer to a 'direct-mode reference' (i) in reference List0 of the co-located slice (if a reference picture from List0 was used for inter-prediction of the co-located macroblock), or (ii) if no List0 reference picture was used, 10 reference List1 of the co-located slice (if a reference picture from List1 was used for inter prediction of the co-located macroblock). The state 106 maps the reference picture from the state 104 to the lowest valued reference index in the current reference List0. The state 106 references the same reference 15 picture that was referenced by the co-located picture in the state 104. The state 108 provides the List0 reference index found in the state 106 (e.g., the reference index is normally generated in response to remapping). The reference index is generally used with the reference list0 of the current slice to determine the second 20 reference picture to be used for inter-prediction.

03-1431  
1496.00341

The state 106 searches the current List0 to determine the lowest valued reference index referring to the same reference picture that was referred by the co-located picture. The step 106 implements a number of operations and data not outlined in the current H.264 specification.

First, a unique identifier for each reference picture is stored. The unique identifier is normally correctly associated from the unique identifier that was used as an inter-reference in the co-located picture. The unique identifier is made available as a potential List0 inter-reference for the current picture. Next, a unique identifier to the actual 'direct-mode reference picture' is stored. Next, a module (or method) searches the current reference List0 for the lowest valued reference index identified by the unique identifier and return the value of that reference index.

A search in the current reference List0 provides the potential for increasing coding efficiency of B-frames and provides flexibility to the encoder to be able to use a truly interpolative direct-mode prediction along with an arbitrary choice for the picture referred by the first reference index (index0) of List0. These two options were mutually exclusive in the previous H.264 implementations discussed in the background section.



03-1431  
1496.00341

Referring to FIG. 2, a diagram 100' is shown in accordance with the present invention. The flow diagram 100' generally comprises a block (or circuit) 112, a block (or circuit) 114, a block (or circuit) 116, a block (or circuit) 118 and a block (or circuit) 119. The block 112 is shown implementing an encoder/decoder signal to construct direct-mode prediction. The block 114 generally sets a co-located picture (e.g., COLPIC) as equal to a value in the List1 [0] (e.g., List1 at index 0). The block 116 generally determines whether the second picture is a picture from either List0 or a List1. The block 118 finds the index in the current List0 that refers to "other picture". The state 119 creates an interprediction by the weighted average of pixels of COLPIC and OTHERPIC.

The present invention may be particularly useful under circumstances when accurate direct-mode prediction is useful. Having a low-overhead/efficient reference to a reference frame other than the two pictures that yield the interpolative direct-mode prediction is desirable. For example, the first entry in List1 of the current B-picture is generally chosen to give a good direct-mode prediction. Higher compression may be achieved by using a picture other than the index 0 entry in List0 of the co-

03-1431  
1496.00341

located picture as the index 0 entry in List0 of the current picture.

Another feature of the present invention is the choice of the order of pictures in the reference lists of the current picture being decoupled and independent from the choice of the order of reference pictures in the co-located picture. Decoupling the reference pictures potentially significantly simplifies the design of an encoder incorporating the present invention. For example, the lists of the co-located picture need not be taken into account when designing the lists for the current picture.

A unique reference index is normally found for the second reference picture for direct-mode predicted blocks. The lowest valued index in List0 is specifically chosen to use the same physical reference frame. The encoder is not unnecessarily constrained to refer to the same physical frame with the same index in List0 of the current picture as was used for the 'direct-mode reference' of the co-located picture. A meaningful and useful direct-mode prediction that effectively finds the current frame as a temporally interpolated intermediate estimate between the co-located picture and a corresponding 'direct-mode L0 reference'. For example, the index 0 of List0 may be chosen to maximize the

03-1431  
1496.00341

coding efficiency of the reference indices rather than to make sure that a good direct-mode prediction is available.

The reference index prediction and context-based coding (e.g., with CABAC entropy-coding) may be improved by using the  
5 LOWEST possible reference index. Reference indices will often be ordered from most-frequent to least-frequent in expected occurrence frequency. The predicted index entropy should be expected to be reduced with the present invention.

The present invention may be implemented in all H.264  
10 compliant decoders. While encoding may be implemented using the techniques described in the background section, the efficiency of such a system is generally reduced when compared with the present invention. Advanced encoders may realize a benefit by exploiting the improved flexibility possible with the use of the present  
15 invention.

A content-addressable-memory (CAM) may provide an efficient hardware structure for implementing the present invention. The present invention may also be implemented in software with a 'for' loop search beginning at index 0 of List0 and  
20 proceeding towards the end of List0. Such a software implementation may exit early from the loop when the desired

03-1431  
1496.00341

reference frame is found. In other implementations, the present invention may be implemented with a variable that holds the lowest reference index in the current List0 that refers to the 'direct-mode L0 reference picture'. Such a variable would be set once  
5 before the decoding of the temporal direct-mode macroblocks of each new slice (e.g., if temporal direct rather than spatial direct is used for the slice) using the software table search 'for' loop method mentioned above.

Referring to FIG. 3, a partial block diagram of an  
10 example implementation of an encoder apparatus 120 is shown. The encoder apparatus 120 may be implemented as a video bitstream encoder apparatus or system. The encoder apparatus 120 generally comprises a circuit 122, a circuit 124, a circuit 126 and a memory 128. The circuit 122 may receive a bitstream or signal (e.g.,  
15 TIN). A bitstream or signal (e.g., TOUT) may be generated by the circuit 126. The memory 128 may hold the List0 and the List1 for each of the reference index values.

The circuit 122 may be implemented as a compression circuit or module. The compression circuit 122 may be operational  
20 to compress the blocks within the signal TIN thereby generating motion vectors. Compression may be determined by a signal (e.g.,

03-1431  
1496.00341

PRED) received from the circuit 124. A signal (e.g., MV) may exchange motion vectors between the compression circuit 122 and the memory 128. During compression, the motion vectors may be written to the memory 128. During reconstruction of a reference block the  
5 motion vectors may be read from the memory 128.

The circuit 124 may be implemented as a code control circuit. The circuit 124 may generate the signal PRED conveying the prediction type used by the macroblocks. The code control circuit 124 may also generate a signal (e.g., CNT). The signal CNT  
10 may provide coding controls to the circuit 126.

The circuit 126 may be implemented as a coding circuit. In one embodiment, the coding circuit 126 may be an entropy coding circuit. The entropy coding circuit 126 may receive the blocks and the associated groups of motion vectors from the compression  
15 circuit 122 via a bitstream or signal (e.g., TBS). The entropy coding circuit 126 may be configured to encode the signal TBS to generate the signal TOUT for transmission and/or storage. In one embodiment, the signal TOUT may be implemented as a Network Abstraction Layer defined by the H.264 standard.

20 The memory 128 may be implemented as an external memory. The memory 128 is generally operational to store the motion vectors

03-1431  
1496.00341

for the blocks while the blocks are being encoded. The memory 128 may be configured to store other data used for encoding the bitstream data. Other types of memories may be implemented to meet the criteria of a particular application.

5           Referring to FIG. 4, a partial block diagram of an example implementation of a decoder apparatus 130 is shown. The decoder apparatus 130 may be implemented as a video bitstream decoder or system. The decoder apparatus 130 generally comprises a circuit 132, a circuit 134, a circuit 136 and a memory 138. The  
10       circuit 132 may receive an input bitstream or signal (e.g., RIN). The circuit 136 may generate an output bitstream or signal (e.g., ROUT).

          The circuit 132 may be implemented as a decoder circuit. In one embodiment, the decoder circuit 132 may be implemented as an  
15       entropy decoder circuit 132. The entropy decoder circuit 132 may be operational to decode the bitstream signal TOUT generated by the entropy coding circuit 126 (e.g., TOUT=RIN). A decoded bitstream or signal (e.g., RBS) may be presented by the entropy decoder circuit 132 to the circuits 134 and 136.

20           The circuit 134 may be implemented as a prediction circuit. The prediction circuit 134 may be operational to

03-1431  
1496.00341

determine if inter or intra prediction has been implemented for the various macroblocks of the pictures in the signal RBS. The prediction circuit 134 may generate a command signal (e.g., CMD) to the circuit 136 indicating the prediction type.

5               The circuit 136 may be implemented as a decompression circuit. The decompression circuit 136 may examine the compressed groups to determine how the motion vectors should be used. The decompression circuit 136 may store the motion vectors from decoded blocks that may be used for inferring motion vectors of co-located  
10 blocks the memory 128 via a signal (e.g., MV). The stored motion vectors may be read from the memory 138 to calculate the motion vectors for B-slice blocks coded under the direct mode (e.g., no associated motion vectors were transmitted in the signal TOUT). The direct mode generally refers to a macroblock or macroblock  
15 partition. The inferred motion vectors may then be used in generating the signal ROUT.

              The memory 138 may be implemented as an external memory. The memory 138 is generally operational to store the motion vectors for the blocks for later use in calculating inferred motion vectors  
20 for the co-located blocks. The memory 138 may be configured to store other data used for decoding the bitstream data. Other types

03-1431  
1496.00341

of memories may be implemented to meet the criteria of a particular application. The memory 138 may hold the List0 and the List1 for each of the reference index values.

5 The present invention may be implemented in decoders of the professional version (PExt) of the H.264 standard that use B-frames, and also for other future extensions of the H.264 standard.

While the invention has been particularly shown and described with reference to the preferred embodiments thereof, it will be understood by those skilled in the art that various changes  
10 in form and details may be made without departing from the spirit and scope of the invention.